

## 91 深入探索多表关联的SQL语句到底是如何执行的？（1）

之前我们已经用很大的篇幅讲完了针对单表的查询SQL语句，通常都会使用哪些执行计划，如何去使用索引去查找数据，想必大家都已经透彻的掌握这些知识了，比如以后在执行计划里看到const、ref、range、index、all以及多索引查询合并的一些字样，都知道具体在磁盘数据层面是如何执行的了

那么今天开始，我们来进入一块极为重要的知识领域，那就是MySQL的多表关联查询SQL语句是如何执行的？

大家都知道，平时一般如果我们仅仅是执行一下单表查询，那都是比较简单的，而且通常你把索引给建好了，让他尽可能走索引，性能都不是什么大问题。

但是往往我们平时基于MySQL做一些系统开发的时候，比较多的是写一些多表关联语句，因为有时候想要查找你需要的数据，不得不借助多表关联的语法去编写SQL语句，才能实现你想要的逻辑和语义，但是往往使用多表关联的时候，你的SQL性能就可能会遇到一些问题。

那么今天开始，我们就一起来看看，这个多表关联SQL语句到底是如何执行的吧。

今天先来给大家讲解一个超级简单，最最基础的多表关联查询的执行原理，假设我们有一个SQL语句是：select \* from t1,t2 where t1.x1=xxx and t1.x2=t2.x2 and t2.x3=xxx

就这么一个SQL语句，大家知道他是什么意思吗？

首先，如果你在FROM子句后直接来了两个表名，这意思就是要针对两个表进行查询了，而且会把两个表的数据给关联起来，假设你要是没有限定什么多表连接条件，那么可能会搞出一个笛卡尔积的东西。

举个例子，假设t1表有10条数据，t2表有5条数据，那么此时select \* from t1,t2，其实会查出来50条数据，因为t1表里的每条数据都会跟t2表里的每条数据连接起来返回给你，那么不就是会查出来 $10 * 5 = 50$ 条数据吗？这就是笛卡尔积

不过通常一般没人会傻到写类似这样的SQL语句，因为查出来这种数据实在是没什么意义。所以通常都会在多表关联语句中的WHERE子句里引入一些关联条件，那么我们回头看看之前的SQL语句里的WHERE子句：where t1.x1=xxx and t1.x2=t2.x2 and t2.x3=xxx

首先呢，t1.x1=xxx，这个可以明确，绝对不是多表关联的连接条件，他是针对t1表的数据筛选条件，本质就是从t1表里筛选一波数据出来再跟t2表做关联的意思。然后t2.x3=xxx，也不是关联条件，他也是针对t2表的筛选条件。

其实真正的关联条件是 $t1.x2=t2.x2$ ，这个条件，意思就是说，必须要让t1表里的每条数据根据自己的x2字段的值去关联上t2表里的某条记录，要求是t1表里这条数据的x2值和t2表里的那条数据的x2字段值是相等的。

举个例子，假设t1表里有1条数据的x2字段的值是265，然后t2表里有2条数据的x2字段的值也是265，那么此时就会把t1表里的那条数据和t2表的2条数据分别关联起来，最终会返回给你两条关联后的数据。

那么基本概念理解清楚了，具体到上面的SQL语句：`select * from t1,t2 where t1.x1=xxx and t1.x2=t2.x2 and t2.x3=xxx`

其实这个SQL执行的过程可能是这样的，首先根据 $t1.x1=xxx$ 这个筛选条件，去t1表里查出来一批数据，此时可能是const、ref，也可能是index或者all，都有可能，具体看你的索引如何建的，他会挑一种执行计划访问方式。

然后假设从t1表里按照 $t1.x1=xxx$ 条件筛选出2条数据，接着对这两条数据，根据每条数据的x2字段的值，以及 $t2.x3=xxx$ 这个条件，去t2表里找x2字段值和x3字段值都匹配的数据，比如说t1表第一条数据的x2字段的值是265，此时就根据 $t2.x2=265$ 和 $t2.x3=xxx$ 这俩条件，找出来一波数据，比如找出来2条吧。

此时就把t1表里x2字段为265的那个数据跟t2表里 $t2.x2=265$ 和 $t2.x3=xxx$ 的两条数据，关联起来，就可以了，t1表里另外一条数据也是如法炮制而已，这就是多表关联最最基本的原理。

记住，他可能是先从一个表里查一波数据，这个表叫做“驱动表”，再根据这波数据去另外一个表里查一波数据进行关联，另外一个表叫做“被驱动表”

**End**